

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

Кафедра «Програмної інженерії та інформаційних
технологій управління»

МЕТОДИЧНІ ВКАЗІВКИ

до лабораторної роботи № 2 за курсом

«Інтелектуальний аналіз даних та видобування знань»

для студентів спеціальностей : спеціальності 121- Інженерія програмного
забезпечення, спеціалізації 121 – 01 - Інженерія програмного забезпечення

Затверджено

редакційно-видавничою радою університету,

протокол № від 2017 року

Харків НТУ «ХПІ» 2017

Методичні вказівки до проведення лабораторних занять за курсом "Інтелектуальний аналіз даних та видобування знань" для студентів спеціальності 121- Інженерія програмного забезпечення, спеціалізації 121 – 01 Інженерія програмного забезпечення \ Уклали – Єршова С. І., Мельник К. В. – Харків: НТУ «ХПІ», 2017. – с.

Укладачі С. І. Єршова, К. В. Мельник

Рецензент О. Ю. Чередніченко

Кафедра програмної інженерії та інформаційних технологій управління

ЗМІСТ

Вступ.....	4
Загальні положення.....	5
Правила виконання і оформлення лабораторних робіт.....	6
Лабораторна робота № 2 «Розробка онтологічного представлення предметної області».....	10
Контрольні питання	20
Література	21

ВСТУП

Методичні вказівки містять завдання для виконання лабораторних робіт для студентів базового напрямку «Програмна інженерія» рівня магістрів.

Мета навчальної дисципліни є опанування студентами теоретичних основ інтелектуального аналізу даних та видобування знань, надання студентам систематизованих знань про основні моделі, методи та технології, які використовуються для інтелектуального аналізу даних та видобування знань; отримання студентами теоретичних знань та практичних навичок роботи з сучасними технологіями, отримання практичних навичок використання програмних засобів розробки онтологій у наукових дослідженнях; формування у студентів аналітичних здібностей, які б дозволили робити обґрунтований вибір методів, засобів та технологій при рішенні задач предметної області.

ЗАГАЛЬНІ ПОЛОЖЕННЯ

Онтологія (від грец. онтос – суще, логос – навчання, поняття) – термін, що визначає вчення про буття, про сутність.

Термін ввів у філософську літературу німецький філософ Р. Гокленіус (1547–1628). До цього онтологія була частиною метафізики, наукою, що не пов’язана з логікою, з науками про природу. Її предмет становить вивчення абстрактних і загальних філософських категорій.

Сьогодні розуміння терміну «онтологія» різне, залежно від контексту і цілей його використання. У роботі [1] виділяють такі аспекти інтерпретації цього терміну:

- 1) онтологія як філософська дисципліна;
- 2) онтологія як неформальна концептуальна система;
- 3) онтологія як формальний погляд на семантику;
- 4) онтологія як специфікація “концептуалізації”;
- 5) онтологія як уявлення концептуальної системи через логічну теорію, що характеризується спеціальними формальними властивостями, її призначенням;
- 6) онтологія як словник;
- 7) онтологія як метарівнева специфікація логічної теорії.

Четверта інтерпретація, запропонована Томасом Грубером. Визначення найчастіше використовується сьогодні в роботах з проектування і дослідження онтологій.

ПРАВИЛА ВИКОНАННЯ І ОФОРМЛЕННЯ ЛАБОРАТОРНИХ РОБІТ

Звіт до лабораторної роботи виконується на одній стороні аркуша білого паперу формату А-4 (210х297 мм) шрифтом Times New Roman розміру 14 через 1,5 міжрядний інтервал. На кожному аркуші повинні бути поля таких розмірів: ліворуч, зверху та знизу - не менше за 20 мм, праворуч - не менше ніж 10 мм. Шрифт друку повинен бути чітким, чорного кольору. Щільність тексту має бути однаковою.

Друкарські помилки, описки чи графічні нечіткості, виявлені під час оформлення роботи, можна виправляти охайним підчищенням чи за допомогою коректора і нанесенням на тому ж місці або між рядками виправленого тексту (фрагмента рисунка) тим самим кольором, яким написаний текст. Заголовки структурних частин роботи «ЗМІСТ», «ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ» (якщо такий є), «ВСТУП», «ХІД ВИКОНАННЯ РОБОТИ», «ВИСНОВКИ», «СПИСОК ДЖЕРЕЛ ІНФОРМАЦІЇ», «ДОДАТКИ» (якщо такі є) друкують великими літерами. Крапка наприкінці заголовка не ставиться.

Текст розділів може складатись з підрозділів. Заголовки підрозділів друкуються маленькими літерами (крім першої великої) з абзацного відступу. Крапка наприкінці заголовка не ставиться.

Якщо підрозділи містять пункти, то заголовки цих пунктів друкують маленькими літерами (крім першої великої) з абзацного відступу. Крапка наприкінці заголовка не ставиться..

Якщо заголовок складається із двох чи більше речень, то їх розділяють крапкою.

Нумерація сторінок, розділів, підрозділів, пунктів, підпунктів, рисунків, таблиць, формул подається арабськими цифрами без знака №.

Нумерація має бути наскрізною, причому першою сторінкою є титульний аркуш, але на ньому номер сторінки не ставиться. Рисунки і Таблиці включаються у загальну нумерацію. Сторінки з додатками і списком літератури

входять у наскрізну нумерацію. Номер сторінки проставляється у правому верхньому кутку сторінки без будь-яких знаків (крапки наприкінці, виділення рисками, дужками тощо).

Підрозділи нумеруються у межах розділу. Номер підрозділу складається із номера розділу та порядкового номера, розділених крапкою. У кінці також ставиться крапка, наприклад «2.1» (перший підрозділ (параграф) другого розділу). Пункти нумеруються арабськими цифрами у межах кожного підрозділу. Номер пункту має складатися із номера розділу, підрозділу і пункту (свого порядкового номера), розділених крапками. У кінці номера пункту ставиться крапка, наприклад, «2.1.3» (третій пункт першого підрозділу другого розділу). Потім у тому ж порядку йде заголовок пункту. Номери підрозділів і пунктів ставляться на їх початку, номер пункту на початку першого рядка абзацу, яким розпочинається відповідний пункт. Цифра номера пункту не повинна виступати за межі абзацу. Якщо пункти міститимуть підпункти, то їх нумерують у межах кожного пункту за такими ж правилами, як пункти.

Ілюстрації (фотографії, схеми, креслення, графіки тощо) називаються рисунками, які нумеруються послідовно у межах розділу або наскрізно арабськими цифрами: номер рисунку складається з двох цифр. Перша - номер розділу, друга - порядковий номер рисунку у розділі, розділених крапкою, наприклад «Рисунок 1.2» (другий рисунок першого розділу), а далі йде назва рисунка. При посиланні на рисунок перший раз необхідно вписувати його повний номер, наприклад, «(рис. 1.2)». Подальші посилання на рисунки виконуються разом із скороченим словом «дивись», наприклад «(див. рис. 1.2)». Рисунки мають розташовуватись одразу після посилання на них у тексті. Якщо на цій сторінці немає місця, то їх необхідно розташувати на наступній сторінці у зручній для ознайомлення формі тобто, щоб для вивчення цього рисунка сторінку можна було б повернути за годинниковою стрілкою. Номер рисунка, його назва і пояснювальні підписи розміщуються послідовно під рисунком.

Таблиці у роботі нумеруються послідовно (за винятком таблиць, поданих у додатках) в межах розділу. У правому верхньому куті пишеться слово «Таблиця» із зазначенням її номера. Номер таблиці складається із номера розділу і порядкового номера таблиці, розділених крапкою, наприклад «Таблиця 1.2» (друга таблиця першого розділу). При посиланні на таблицю вказують її повний номер, а слово «Таблиця» пишуть скорочено, наприклад «(табл. 1.2)». Якщо у роботі тільки одна таблиця, то її не нумерують і слово «Таблиця» не пишуть.

Під словом «Таблиця» розміщуються заголовок таблиці симетрично до форми таблиці. Слово «Таблиця» і заголовок починаються з великої букви. Назва не підкреслюється.

Таблиці потрібно розташовувати після першої згадки про неї у тексті так, щоб її зручно було розглядати без повороту або з поворотом за годинниковою стрілкою. Якщо на цій сторінці немає місця, то їх необхідно розташувати на наступній сторінці у зручній для ознайомлення формі, тобто, щоб для вивчення таблиці сторінку можна було б повернути за годинниковою стрілкою.

При переносі таблиці на наступну сторінку нумерацію колонок необхідно повторити, а над нею розмістити слова «Продовження табл.» із зазначенням її номера, наприклад «Продовження табл. 1.2».

Таблицю з великою кількістю рядків можна переносити на інший аркуш, при цьому назву розміщують тільки над її першою частиною. Таблицю з великою кількістю граф можна ділити на частини і розміщувати одну частину під іншою у межах одної сторінки. Якщо рядки або графи таблиці виходять за формат сторінки, то у першому випадку в кожній частині таблиці повторюють її головку.

Висота рядків повинна бути не меншою за 8 мм.

Формули в роботі (якщо їх більше від однієї) нумерують у межах розділу. Номер формули складається із номера розділу і порядкового номера формули у розділі, між якими ставиться крапка. Номер формули пишеться біля правого

краю аркуша на рівні відповідної формули (якщо формула велика, то на рівні нижнього рядка формули до якої він стосується) в круглих дужках, наприклад (2.2) (друга формула другого розділу). Пояснення позначень символів чи числових коефіцієнтів наводиться безпосередньо під формулою у тій послідовності, у якій вони подані у формулі. Це пояснення подається з нового рядка, починаючи зі слова "де", двокрапка після якого не ставиться. Значення кожного символу чи числового коефіцієнта подається з нового рядка. Розмірність одного і того самого параметра у межах роботи має бути однаковою. При посиланні у тексті на формулу, необхідно подати її повний номер у дужках, наприклад: «У формулі (1.2)».

Вище і нижче від кожної формули потрібно залишити не менше ніж один вільний рядок.

До списку джерел інформації входять усі використані джерела, які розташовуються за алфавітом або у послідовності посилань на них у тексті. Інформація про видання (монографії, підручники, довідники тощо) має містити: прізвище та ініціали автора, назву книги, місце видання, видавництво і рік видання, обсяг у сторінках.

Прізвище автора подається у називному відмінку. Якщо є два, три чи чотири автори, то їх прізвища з ініціалами подають у тій послідовності, у якій вони надруковані у книзі: перед прізвищем наступного автора ставлять кому. Якщо більше ніж чотири автори, вказують прізвища з ініціалами тільки перших трьох, а далі пишуть слова "та інші". Місце видання приводиться повністю у називному відмінку, можна скорочувати назви тільки : Київ (К.).

Варіант задається за порядковим номером студента у журналі.

ЛАБОРАТОРНА РОБОТА № 2

«РОЗРОБКА ОНТОЛОГІЧНОГО ПРЕДСТАВЛЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ»

Мета роботи: Розробка онтології в Protege для заданої предметної області.

Завдання:

1. Познайомитись з середовищем Protege.
2. Розробити онтологію у форматі OWL.
3. Розробити ієрархію класів в онтології згідно результатів попередньої лабораторної роботи.
4. Візуалізувати онтологію у вигляді графу.
5. Розробити властивості класів онтології, які повинні описувати характеристики класів та відносини між класами.
6. Створити екземпляри класів.
7. Заповнити онтологію можливими прецедентами (25-40 прецедентів).

Довідковий матеріал до лабораторної роботи:

Для розробки онтології обраної предметної області необхідно використати редактор онтологій Protege. Це вільний редактор, який дозволяє моделювати предметну область у вигляді онтології та розробляти базу знань.

Завантажити останню версію редактора Protege можна з офіційного сайту:
<https://protege.stanford.edu/>.

Для коректної роботи цієї програми необхідно перевірити наявність Java на комп'ютері. Якщо Java не встановлена, тоді треба перейти за посиланням https://www.java.com/ru/download/help/download_options.xml та виконати всі інструкції.

Хід роботи

1 Для того, щоб створити нову онтологію, треба запустити редактор, ввести назву онтології та анотацію щодо предметної області (рис. 2.1).

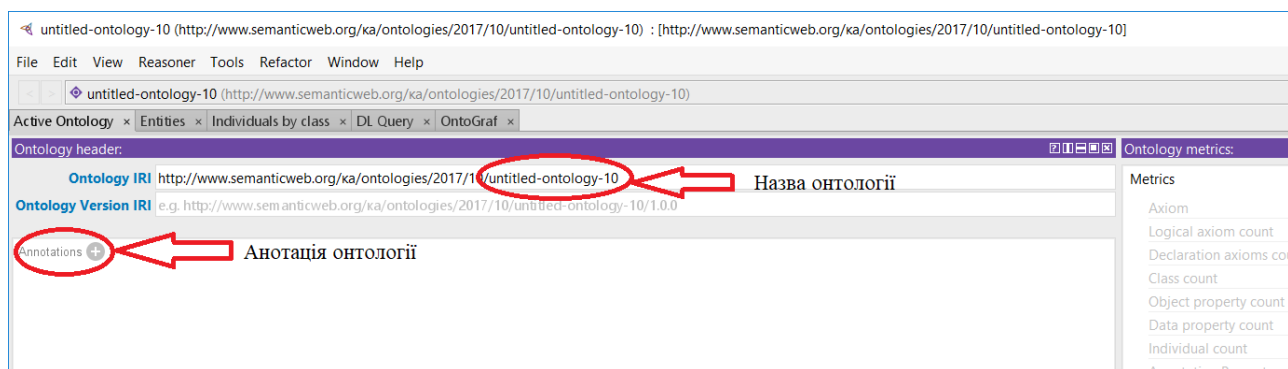


Рисунок 2.1 – Головне вікно Protege

Анотація повинна містити стислу інформацію щодо мети створення онтології (рис.2.2) .

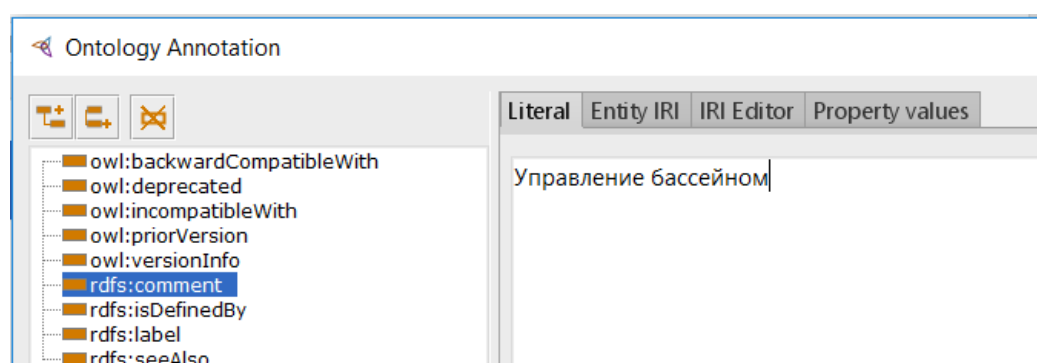


Рисунок 2.2 – Створення анотації для онтології «Управління басейном»

2 Після проведених дій необхідно зберегти онтологію у форматі owl (рис.2.3).

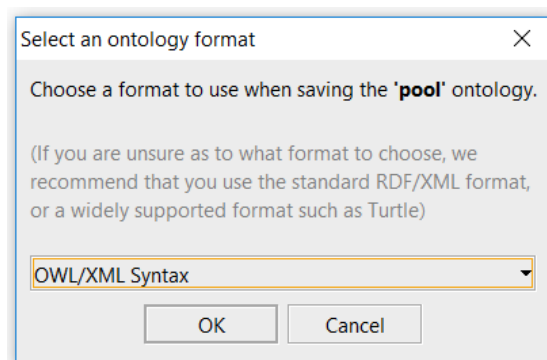


Рисунок 2.3 – Збереження онтології

3 Щоб створити клас в Protege, необхідно перейти на вкладку Classes (Entities→Classes), обрати клас «owl: Thing» та вибрати в меню «AddSubclass», або натиснути кнопку створення класу (рис.2.4). Якщо в редакторі вкладка Classes не відображається, потрібно на панелі інструментів вибрати Window → Tabs → Classviews.

В OWL базовим класом, на основі якого створюються класи онтології, є клас «owl: Thing». Решта класів по відношенню до нього є дочірніми підкласами (англ. Subclass). Класи одного рівня ієрархії в Protege називаються спорідненими (англ. Sibling Class).

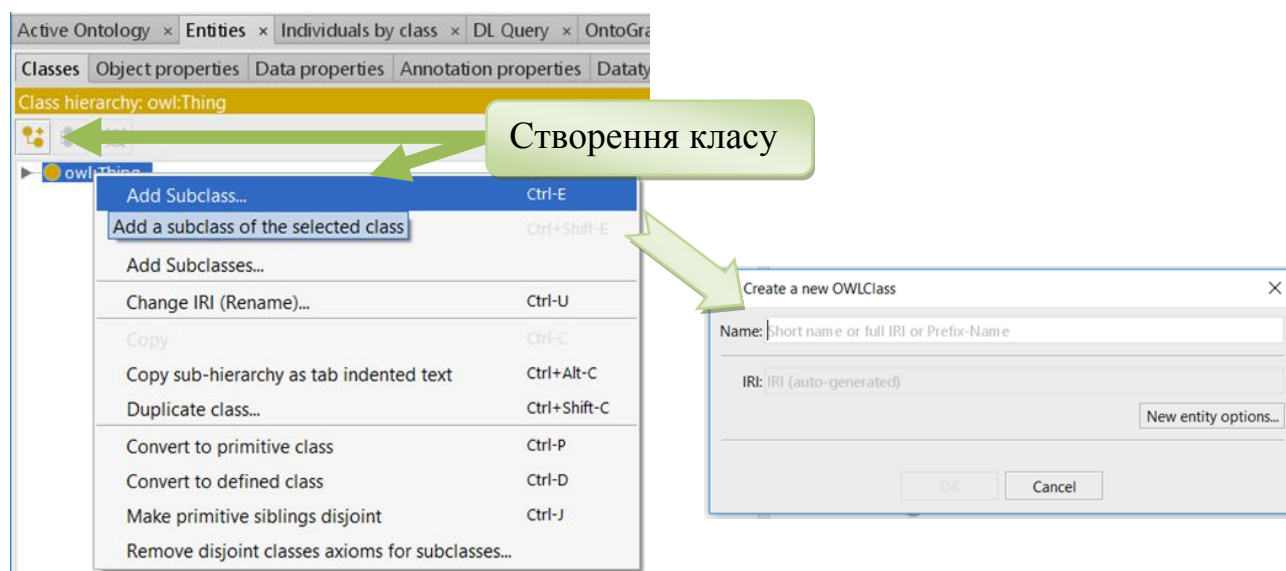


Рисунок 2.4 – Створення нового класу

Для кожного класу треба створити анотацію на вкладці «ClassAnnotations», яка пояснює за що відповідає створений клас. На вкладці «Class Usage» для класу, виділеного в ієрархії, відображається його зв'язок з батьківськими і дочірніми класами, його властивості, екземпляри і т.п. У правій нижній панелі «Description» необхідно вказати додаткові характеристики класу. Наприклад, його приналежність іншим класам «Subclassof» або неможливість приналежності екземплярам інших класів «DisjointWith».

В результаті аналізу управління басейном, який був проведений при виконанні попередньої роботи, були виділені наступні сутності:

- Condition - вхідні умови для управління басейном:
 - Temp_env - температура навколишнього середовища;
 - Temp_water - температура басейну;
 - Depth - рівень води в басейні.
- Action - можливі дії для нормалізації умов перебування в басейні:
 - Heater - включити/виключити підігрівач води;
 - Stopper - відкрити/закрити кришку зливу води;
 - Water-tap - включити/виключити подачу води.

Для опису можливих ситуацій також треба зробити клас Case, який буде містити інформацію щодо прийняття рішень при управлінні басейном.

В результаті була сформована наступна ієрархія класів (рис.2.5):

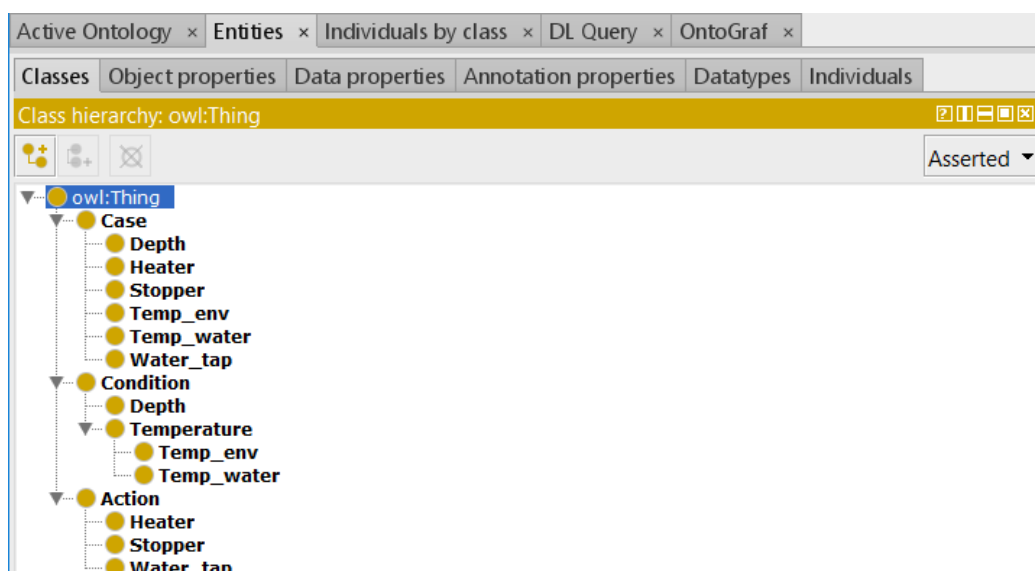


Рисунок 2.5 – Ієрархія класів

4 Візуалізацію розробленої онтології у вигляді графа можна побачити на вкладці «OntoGraf» (рис.2.6):

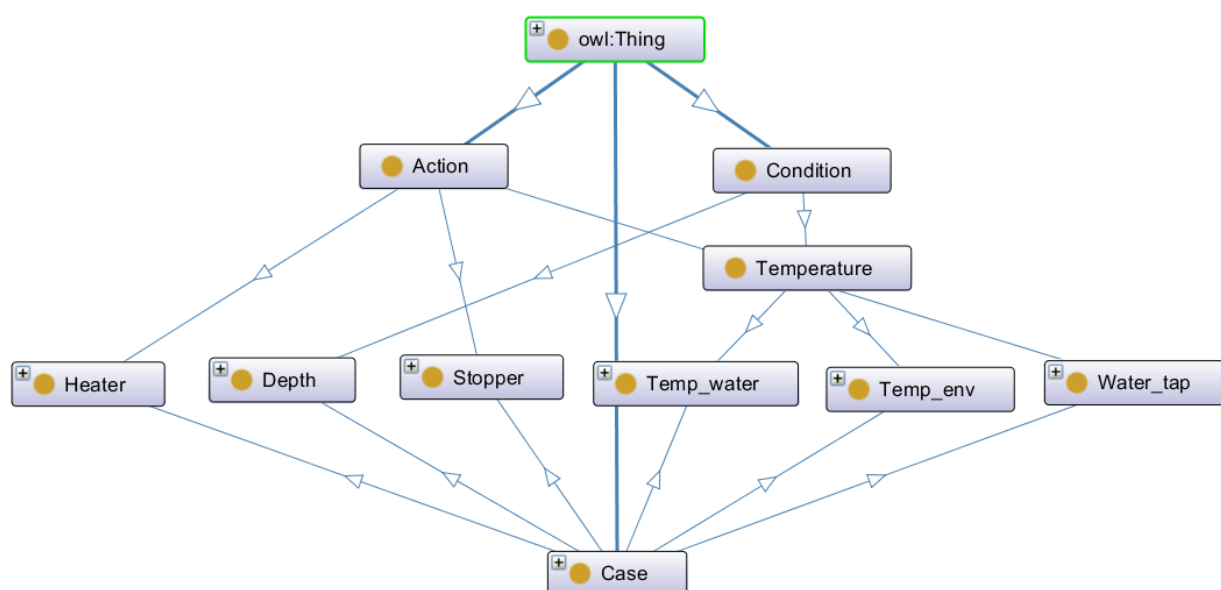


Рисунок 2.6 – Графічна інтерпретація онтології

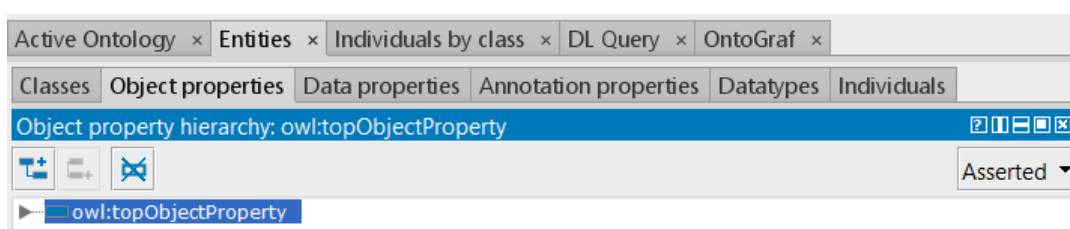
Коли до онтології будуть добавлені властивості класів та екземпляри класів, то графічний вигляд онтології зміниться за рахунок додаткової інформації.

5 Щоб створити нову властивість або редагувати існуючу, спочатку треба визначити тип властивості:

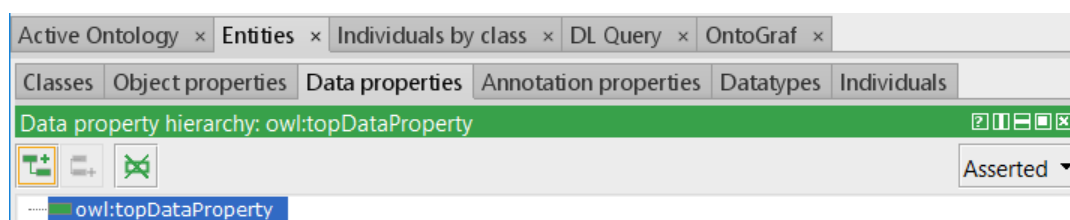
- властивість, що відповідає за відносини між екземплярами класів знаходиться на вкладці «Object Properties» (рис.2.7 а);

- у випадку, коли властивість відповідає за характеристики класів, необхідно вибрати вкладку «Data Properties» (рис.2.7 б).

Процес створення нової властивості аналогічний процесу створення нового класу.



а)



б)

Рисунок 2.7 – Створення властивості: а) властивість-відносини

б) властивість-характеристика

Ієрархія властивостей-відносини представлена на рис. 2.8.

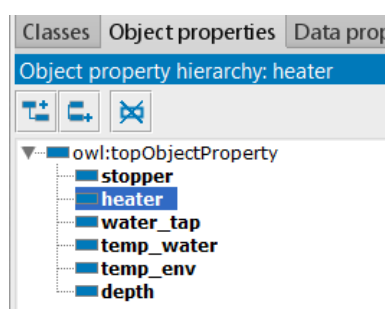


Рисунок 2.8 – Ієрархія властивостей

Для кожної властивості необхідно вказати значення. Для цього потрібно обрати властивість, потім в правій частині вікна натиснути на «Ranges» чи «Domain».

Домен (англ. Domain) вказує, екземпляри яких класів при використанні даної властивості будуть виступати в якості суб'єктів, а діапазон (англ. Range) - в якості об'єктів.

У вікні, що з'явилося обирають властивість та клас, який описується за допомогою цієї властивості (рис. 2.9).

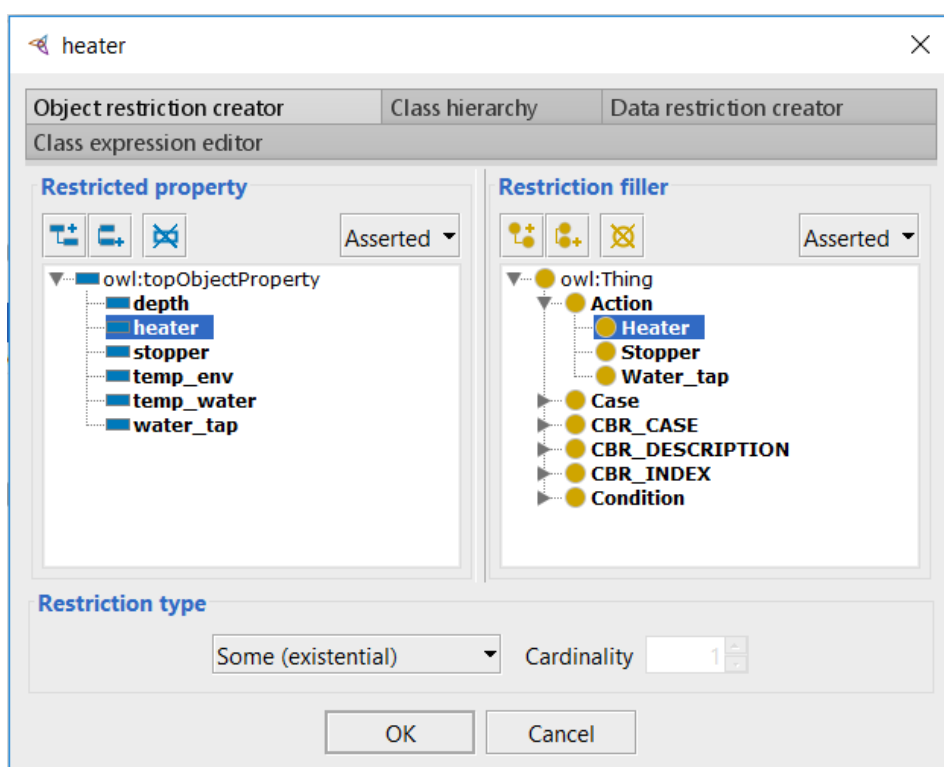


Рисунок 2.9 – Створення зв'язку між класом та властивістю

Результат цієї дії представлений на рис. 2.10.

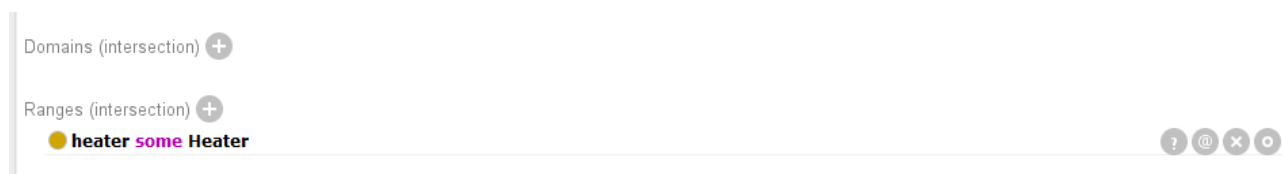


Рисунок 2.10 – Зв'язок між класом та властивістю

Приклад фрагменту ієрархії властивостей даних та процес створення значень приведені на рис. 2.11-2.13. Для даного типу властивості домен та діапазон мають наступні визначення:

- домен (англ. Domain) вказує, для примірників яких класів дане властивість може бути використано;
- діапазон (англ. Range) задає область допустимих значень (тип даних і обмеження), які можна вказувати для даного властивості екземпляру класу; тип даних вибирається зі стандартних типів XML.

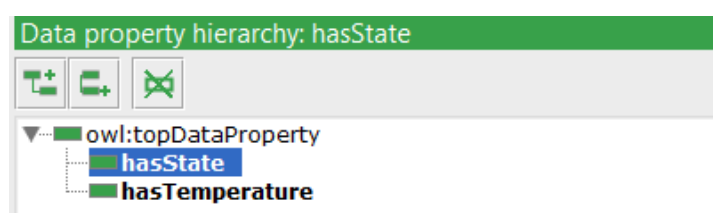


Рисунок 2.11 – Фрагмент ієрархії властивостей даних

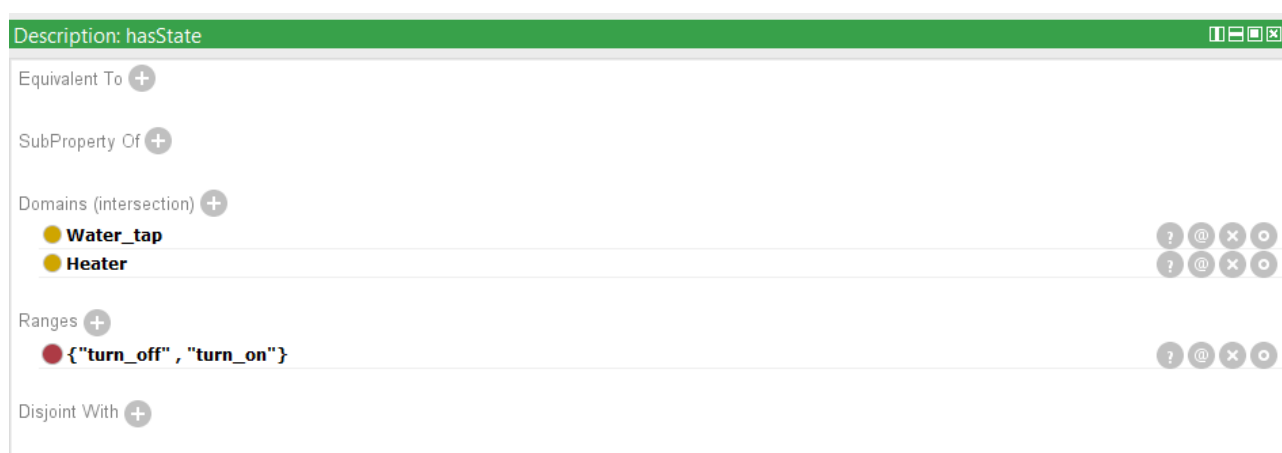


Рисунок 2.12 – Заповнення значеннями властивості «hasState»

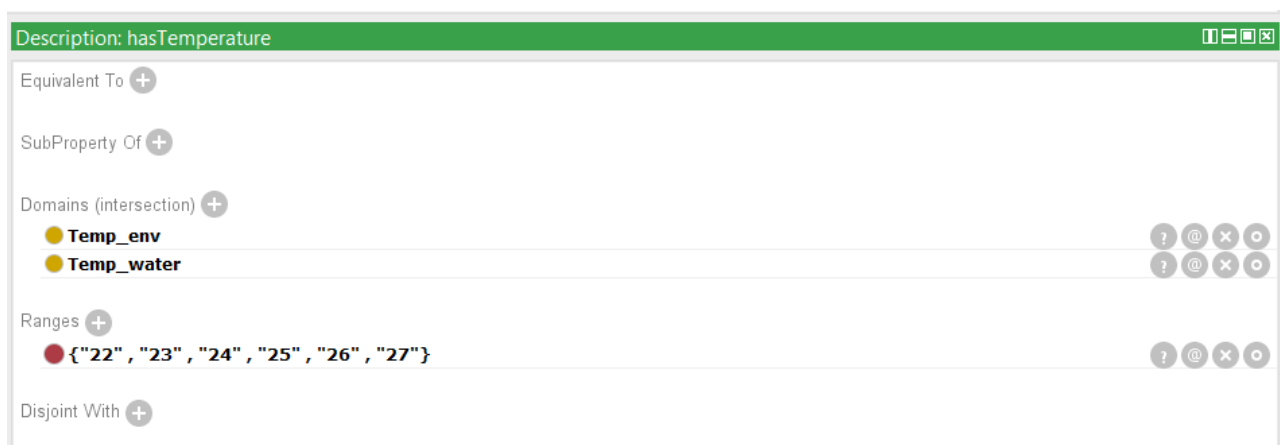



Рисунок 2.13 – Заповнення значеннями властивості «hasTemperature»

6 Процес створення екземплярів класів: необхідно обрати вкладку «Individuals», а потім кнопку , яка відповідає за створення нового екземпляру. Приклад результату роботи приведений на рис.2.14

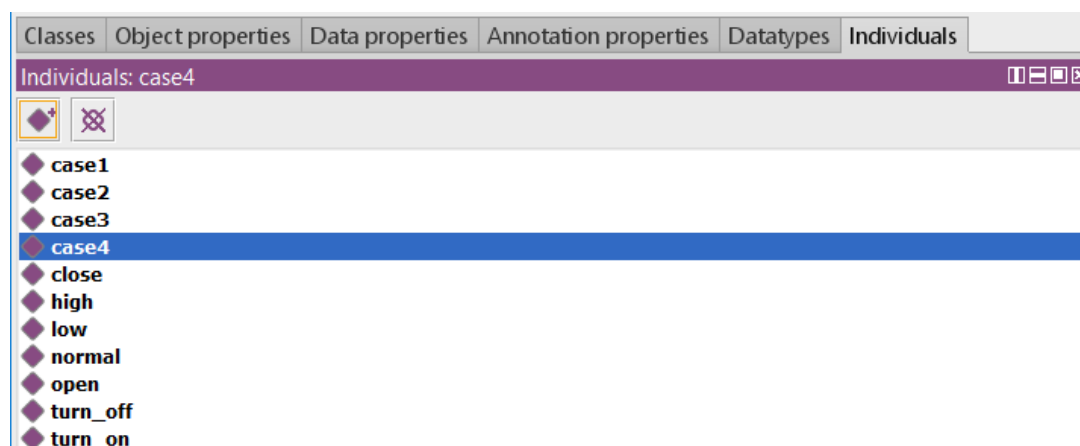


Рисунок 2.14 – Екземпляри класів

7 Для того, щоб заповнити онтологію можливими прецедентами для обраної предметної області, необхідно створити близько 30-40 екземплярів класу «Case». Це статистичні дані, які показують при яких вхідних умовах було використано те чи інше вирішення ситуації. Приклад створення прецеденту показаний на рис. 2.15.

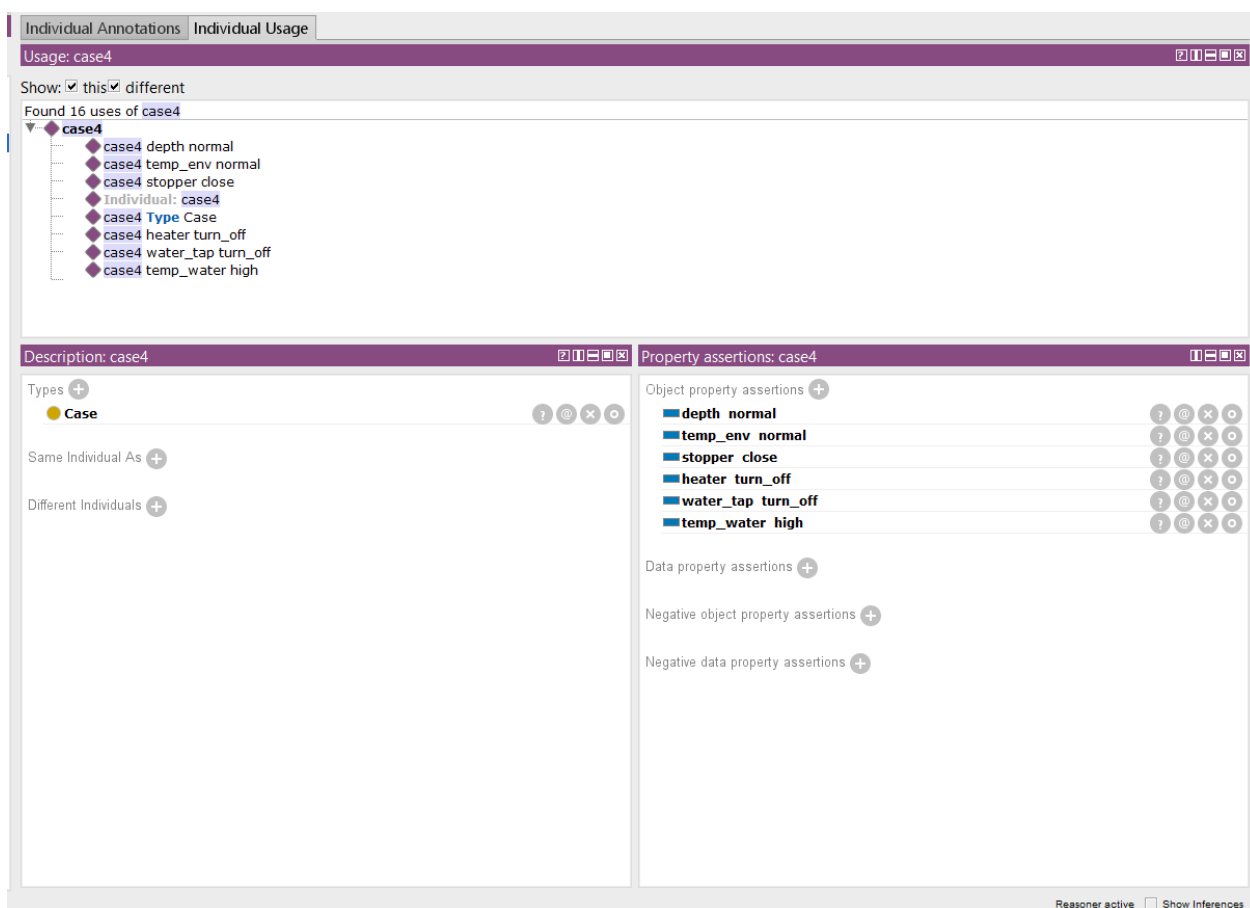


Рисунок 2.15 – Процес створення певного прецеденту

КОНТРОЛЬНІ ПИТАННЯ

1. Які ви знаєте засоби для розробки онтологій?
2. Які версії мови OWL ви знаєте?
3. Як відбувається подання класів в OWL?
4. Методика розробка онтології засобами Protege.
5. Як створити властивість класу у Protege?
6. Як розроблюється структура класів в OWL?
7. Які характеристики мають OWL властивості?
8. Як можна побудувати ієрархію класів?
9. Чи доречно мати імена класів як в однині так і в множині?
10. Як визначається потужність властивостей? В чому різниця розробки онтології та проектування класів і зв'язків в ООП?
11. Призначення фацету властивостей.
12. Охарактеризуйте властивості об'єктів і даних. Наведіть приклади.

ЛІТЕРАТУРА

1. Литвин В.В. Технології менеджменту знань: навч. посібник/ В.В. Литвин. – Львів: Видавництво Львівської політехніки, 2010.-260 с.
2. Башмаков
3. Верников Г. Стандарт онтологического исследования IDEF5-
[www.vernikov.ru](http://citforum.ru/cfin/idef/idef5.shtml)<http://citforum.ru/cfin/idef/idef5.shtml>, 11.11.2017
4. Д.И. Муромцев. Онтологический инжиниринг знаний в системе Protégé. – СПб: СПб ГУ ИТМО, 2007. – 62 с.